第二章

最小二乘估计（一）

Outline

线性回归模型的经典假设和最小二乘估计

线性回归模型的经典假设

1. Assumption 1.1 (linearity):

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i, \quad (i = 1, 2, \ldots, N),$$

where β's are unknown parameters to be estimated, and ε_i is the unobserved error term.

矩阵形式：定义 $K \times 1$ 维向量

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$

则有,

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i. \tag{1}$$

再定义

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_N' \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1K} \\ x_{21} & \cdots & x_{2K} \\ \vdots & \cdots & \vdots \\ x_{N1} & \cdots & x_{NK} \end{bmatrix}$$

则有，

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

2. Assumption 1.2 (strict exogeneity):

$$\mathbb{E}(\varepsilon_i|\mathbf{X}) = 0, \quad (i = 1, 2, \ldots, N)$$

Remark

- The conditional mean $\mathbb{E}(\varepsilon_i|\mathbf{X})$ is in general a nonlinear function of $\mathbf{X}$. The strict exogeneity assumption says that this function is a constant of value zero.
- Assuming this constant to be zero is not restrictive if the regressors include an intercept.
- The strict exogeneity assumption has several implications:
  - $\mathbb{E}(\varepsilon_i) = 0, \quad (i = 1, 2, \ldots, N)$
  - $\mathbb{E}(x_{jk}\varepsilon_i) = 0, \quad (i, j = 1, 2, \ldots, N; k = 1, \ldots, K)$
  - $\mathrm{Cov}(\varepsilon_i, x_{jk}) = 0$
- For most time-series models, strict exogeneity is not satisfied.

3. Assumption 1.3 (no multicollinearity): The rank of the $N \times K$ data matrix, $\mathbf{X}$, is K with probability 1.
4. Assumption 1.4 (spherical error variance):

$$\begin{aligned}\mathbb{E}(\varepsilon_i^2|\mathbf{X}) &= \sigma^2 > 0, \quad (i = 1, 2, \ldots, N) &(2)\\ \mathbb{E}(\varepsilon_i\varepsilon_j|\mathbf{X}) &= 0, \quad (i, j = 1, 2, \ldots, N; i \neq j). &(3)\end{aligned}$$

Assumption 1.4 can be written compactly as

$$\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}) = \mathrm{Var}(\boldsymbol{\varepsilon}|\mathbf{X}) = \sigma^2\mathbf{I}_N. \tag{4}$$

Remark

- Assumption 1.3 implies that $N \geq K$.
- If $K >> N$，可以使用 LASSO, Ridge Regression or Elastic Net.

同方差 or 异方差

异方差是常态，同方差是特例！

由 Regression CEF Theorem，回归是对 CEF 的近似，即使数据真的是同方差，但 CEF 是非线性的，回归模型的误差项也将是异方差的。具体地，

$$
\begin{aligned}
\mathbb{E}(\varepsilon_i^2 \mid \mathbf{x}_i) &= \mathbb{E}[(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2 \mid \mathbf{x}_i] \\
&= \mathbb{E}\left\{[y_i - \mathbb{E}(y_i \mid \mathbf{x}_i) + \mathbb{E}(y_i \mid \mathbf{x}_i) - \mathbf{x}_i'\boldsymbol{\beta}]^2 \mid \mathbf{x}_i\right\} \\
&= \mathrm{Var}(y_i \mid \mathbf{x}_i) + [\mathbb{E}(y_i \mid \mathbf{x}_i) - \mathbf{x}_i'\boldsymbol{\beta}]^2 .
\end{aligned}
$$

即使 $\mathrm{Var}(y_i \mid \mathbf{x}_i) = \sigma^2$，如果 CEF 是非线性的，第二项将不等于 0，而是 $\mathbf{x}_i$ 的函数。

即使 CEF 是线性的，误差项也可能是异方差的。例如，考虑线性概率模型：

$$y_i = \alpha + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad \mathbb{E}(\varepsilon_i \mid \mathbf{x}_i) = 0,$$

其中，y_i 是二值的。此模型的误差项是异方差的：

$$\begin{aligned}
\mathbb{E}(\varepsilon_i^2 \mid \mathbf{x}_i) &= \mathbb{E}\left\{[y_i - \mathbb{E}(y_i \mid \mathbf{x}_i)]^2 \mid \mathbf{x}_i\right\} \\
&= \mathbb{E}\left\{[y_i - P(y_i = 1 \mid \mathbf{x}_i)]^2 \mid \mathbf{x}_i\right\} \\
&= [1 - P(y_i = 1 \mid \mathbf{x}_i)]^2 P(y_i = 1 \mid \mathbf{x}_i) \\
&\quad + [P(y_i = 1 \mid \mathbf{x}_i)]^2 [1 - P(y_i = 1 \mid \mathbf{x}_i)] \\
&= P(y_i = 1 \mid \mathbf{x}_i)[1 - P(y_i = 1 \mid \mathbf{x}_i)].
\end{aligned}$$

最小二乘估计

对 β 的 hypothetical value $\tilde{\beta}$ 定义第 i 个观测值的残差（residual）:

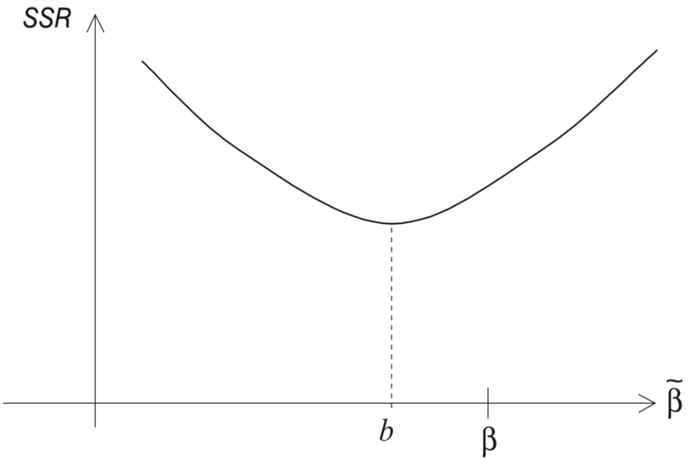
$$y_i - \mathbf{x}_i' \tilde{\boldsymbol{\beta}}.$$

则，残差平方和（Sum of Squared Residuals, SSR）可定义为：

$$SSR(\tilde{\boldsymbol{\beta}}) \equiv \sum_{i=1}^{N} (y_i - \mathbf{x}_i' \tilde{\boldsymbol{\beta}})^2 = (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}). \tag{5}$$

β 的最小二乘估计量 $\boldsymbol{b}$ 通过最小化残差平方和得到，即，

$$\boldsymbol{b} \equiv \arg\min_{\tilde{\boldsymbol{\beta}}} SSR(\tilde{\boldsymbol{\beta}}).$$


SSR
b
β
β̃

Normal Equations: $\mathbf{X}'\mathbf{X}\boldsymbol{b} = \mathbf{X}'\mathbf{y}$.

Proof.

Write

$$\begin{aligned} SSR(\tilde{\boldsymbol{\beta}}) &= (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) = (\mathbf{y}' - \tilde{\boldsymbol{\beta}}'\mathbf{X}')(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\tilde{\boldsymbol{\beta}} \\ &\equiv \mathbf{y}'\mathbf{y} - 2\boldsymbol{a}'\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\beta}}'\mathbf{A}\tilde{\boldsymbol{\beta}}. \end{aligned}$$

Therefore,

$$\frac{\partial SSR(\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}} = -2\boldsymbol{a} + 2\mathbf{A}\tilde{\boldsymbol{\beta}}$$

Setting this equal to zero, we have

$$\mathbf{X}'\mathbf{X}\boldsymbol{b} = \mathbf{X}'\mathbf{y}.$$

□

由 Normal Equations 有：

$$\mathbf{X}'(\mathbf{y}-\mathbf{X}\boldsymbol{b})=\mathbf{X}'\boldsymbol{e}=\mathbf{0},$$

其中，$\boldsymbol{e}=\mathbf{y}-\mathbf{X}\boldsymbol{b}$，称为 OLS residuals.

Remark

- Recall that the FOC of the Best Linear Predictor is $\mathbb{E}(\mathbf{x}e)=\mathbf{0}$. Therefore, normal equations can be interpreted as the sample analogue of this orthogonality conditions.
- 由 Assumption 1.3 容易证明 $\mathbf{X}'\mathbf{X}$ 是正定阵，即二阶海塞矩阵正定，所以 b 是最小化的解。

Useful Definitions:

- 拟合值 (fitted value): $\hat{y}_i \equiv \mathbf{x}_i'\boldsymbol{b}$, $\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{b}$. (Therefore, $\boldsymbol{e} = \mathbf{y} - \hat{\mathbf{y}}$)
- 投影矩阵 (projection matrix): $\mathbf{P} \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.
- Annihilator: $\mathbf{M} \equiv \mathbf{I}_N - \mathbf{P}$.
- Sum of squared OLS residuals: $SSR = \boldsymbol{e}'\boldsymbol{e} = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$.

Useful Facts about $\mathbf{P}$ and $\mathbf{M}$:

- $\mathbf{P}$ 和 $\mathbf{M}$ 是对称幂等阵
- $\mathbf{PX} = \mathbf{X}$.
- $\mathbf{MX} = \mathbf{0}$.

Frisch-Waugh Theorem

考虑如下回归形式：

$$\mathbf{y} = \begin{bmatrix} \mathbf{X}_r & \mathbf{X}_s \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_r \\ \boldsymbol{\beta}_s \end{bmatrix} + \boldsymbol{\varepsilon}.$$

Frisch-Waugh 定理指出，为了分别得到 $\boldsymbol{\beta}_r$ 和 $\boldsymbol{\beta}_s$ 的估计量，可以按照以下步骤操作：

1. 将 $\mathbf{y}$ 对 $\mathbf{X}_r$ 回归，得到残差向量 $\mathbf{y}^*$;
2. 将 $\mathbf{X}_s$ 的每列分别对 $\mathbf{X}_r$ 回归，所得残差向量组成矩阵 $\mathbf{X}_s^*$;
3. $\mathbf{y}^*$ 对 $\mathbf{X}_s^*$ 回归，得到 $\boldsymbol{\beta}_s$ 的 OLS 估计量 $\boldsymbol{b}_s$;
4. 以 $\mathbf{y} - \mathbf{X}_s\boldsymbol{b}_s$ 为被解释变量对 $\mathbf{X}_r$ 进行回归，得到 $\boldsymbol{\beta}_r$ 的 OLS 估计量 $\boldsymbol{b}_r$.

Proof.

Note that

$$\begin{aligned}
\begin{bmatrix} \boldsymbol{b}_r \\ \boldsymbol{b}_s \end{bmatrix} &= \begin{bmatrix} \mathbf{X}_r'\mathbf{X}_r & \mathbf{X}_r'\mathbf{X}_s \\ \mathbf{X}_s'\mathbf{X}_r & \mathbf{X}_s'\mathbf{X}_s \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_r' \\ \mathbf{X}_s' \end{bmatrix} \mathbf{y} \\
&= \begin{bmatrix} \boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{F}_2\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{F}_2 \\ -\boldsymbol{F}_2\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{F}_2 \end{bmatrix} \begin{bmatrix} \mathbf{X}_r'\mathbf{y} \\ \mathbf{X}_s'\mathbf{y} \end{bmatrix}
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{A}_{11} &= \mathbf{X}_r'\mathbf{X}_r, \quad \boldsymbol{A}_{12} = \mathbf{X}_r'\mathbf{X}_s, \quad \boldsymbol{A}_{21} = \mathbf{X}_s'\mathbf{X}_r, \quad \boldsymbol{A}_{22} = \mathbf{X}_s'\mathbf{X}_s, \\
\boldsymbol{F}_2 &= \left(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\right)^{-1} = \left(\mathbf{X}_s'\mathrm{M}_r\mathbf{X}_s\right)^{-1}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\boldsymbol{b}_s &= -\boldsymbol{F}_2\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\mathbf{X}_r'\mathbf{y} + \boldsymbol{F}_2\mathbf{X}_s'\mathbf{y} = \boldsymbol{F}_2\mathbf{X}_s'\mathrm{M}_r\mathbf{y} \\
&= \left(\mathbf{X}_s'\mathrm{M}_r\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\mathrm{M}_r\mathbf{y} = \left[(\mathrm{M}_r\mathbf{X}_s)'(\mathrm{M}_r\mathbf{X}_s)\right]^{-1}(\mathrm{M}_r\mathbf{X}_s)'(\mathrm{M}_r\mathbf{y}).
\end{aligned}$$

On the other hand,

$$\mathbf{X}_r'\mathbf{X}_r\boldsymbol{b}_r + \mathbf{X}_r'\mathbf{X}_s\boldsymbol{b}_s = \mathbf{X}_r'\mathbf{y} \quad \Rightarrow \quad \boldsymbol{b}_r = \left(\mathbf{X}_r'\mathbf{X}_r\right)^{-1}\mathbf{X}_r(\mathbf{y} - \mathbf{X}_s\boldsymbol{b}_s).$$

□

拟合优度（Goodness of fit）

Uncentered R^2: $R^2_{uc} \equiv 1 - \boldsymbol{e}'\boldsymbol{e}/\mathbf{y}'\mathbf{y}$.

被解释变量 $\mathbf{y}$ 的变化可以用 $\sum y_i^2 = \mathbf{y}'\mathbf{y}$ 度量。根据 $\boldsymbol{e} = \mathbf{y} - \hat{\mathbf{y}}$，容易证明：

$$
\begin{aligned}
\mathbf{y}'\mathbf{y} &= (\hat{\mathbf{y}} + \boldsymbol{e})'(\hat{\mathbf{y}} + \boldsymbol{e}) = \hat{\mathbf{y}}'\hat{\mathbf{y}} + 2\hat{\mathbf{y}}'\boldsymbol{e} + \boldsymbol{e}'\boldsymbol{e} \\
&= \hat{\mathbf{y}}'\hat{\mathbf{y}} + 2\boldsymbol{b}'\mathbf{X}'\boldsymbol{e} + \boldsymbol{e}'\boldsymbol{e} \\
&= \hat{\mathbf{y}}'\hat{\mathbf{y}} + \boldsymbol{e}'\boldsymbol{e}.
\end{aligned}
$$

所以，

$$
R^2_{uc} = \frac{\hat{\mathbf{y}}'\hat{\mathbf{y}}}{\mathbf{y}'\mathbf{y}}.
$$

Thus, the uncentered R^2 has the interpretation of the fraction of the variation of the dependent variable that is attributable to the variation in the explanatory variables.

(Centered) R^2: R^2 又称为 coefficient of determination, 其定义如下:

$$R^2 \equiv 1 - \frac{\sum_{i=1}^N e_i^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad \text{with } \bar{y} \equiv \frac{1}{N}\sum_{i=1}^N y_i.$$

如果解释变量中含有常数项，可以证明:

$$\sum_{i=1}^N (y_i - \bar{y})^2 = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^N e_i^2. \tag{6}$$

由 (6), R^2 可以表示成:

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2}.$$

Proof.

为了证明 (6)，将等式左边写成向量内积形式：

$$\sum_{i=1}^{N}(y_i-\bar{y})^2=(\mathbf{y}-\mathbf{i}\bar{y})'(\mathbf{y}-\mathbf{i}\bar{y}),$$

其中，$\mathbf{i}$ 表示 $N\times 1$ 维由 1 构成的列向量。注意到：

$$\begin{aligned}
& (\mathbf{y}-\mathbf{i}\bar{y})'(\mathbf{y}-\mathbf{i}\bar{y}) \\
= & (\mathbf{y}-\hat{\mathbf{y}}+\hat{\mathbf{y}}-\mathbf{i}\bar{y})'(\mathbf{y}-\hat{\mathbf{y}}+\hat{\mathbf{y}}-\mathbf{i}\bar{y}) \\
= & (\hat{\mathbf{y}}-\mathbf{i}\bar{y})'(\hat{\mathbf{y}}-\mathbf{i}\bar{y})+(\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}})+2(\hat{\mathbf{y}}-\mathbf{i}\bar{y})'\underbrace{(\mathbf{y}-\hat{\mathbf{y}})}_{\boldsymbol{e}}.
\end{aligned}$$

若回归中含有截距项，则有：$\mathbf{i}'\boldsymbol{e}=0$. 又因为 $\hat{\mathbf{y}}'\boldsymbol{e}=0$，(6) 式得证。 □

Remark:

- R^2 is nondecreasing when a variable is added to a regression.
- Adjusted R^2: $\bar{R}^2 \equiv 1 - \dfrac{\boldsymbol{e}'\boldsymbol{e}/(N-K)}{\left[\sum_{i=1}^{N}(y_i - \bar{y})^2\right]/(N-1)}$.

最小二乘估计的小样本性质

最小二乘估计的小样本性质

Theorem (Finite-sample properties of the OLS estimator $\boldsymbol{b}$)

(a) *(Unbiasedness) Under Assumptions 1.1-1.3,* $\mathbb{E}(\boldsymbol{b}|\mathbf{X}) = \boldsymbol{\beta}$.

(b) *Under Assumptions 1.1-1.4,* $\mathrm{Var}(\boldsymbol{b}|\mathbf{X}) = \sigma^2 \cdot (\mathbf{X}'\mathbf{X})^{-1}$.

(c) *(Gauss-Markov Theorem) Under Assumptions 1.1-1.4, the OLS estimator is efficient in the class of linear unbiased estimators. That is, for any unbiased estimator* $\hat{\boldsymbol{\beta}}$ *that is linear in* $\mathbf{y}$, $\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) \geq \mathrm{Var}(\boldsymbol{b}|\mathbf{X})$ *in the matrix sense.*[a]

(d) *Under Assumptions 1.1-1.4,* $\mathrm{Cov}(\boldsymbol{b}, \boldsymbol{e}|\mathbf{X}) = \mathbf{0}$, *where* $\boldsymbol{e} \equiv \mathbf{y} - \mathbf{X}\boldsymbol{b}$.

---

[a] For this reason the OLS estimator is called the Best Linear Unbiased Estimator (BLUE).

Proof.

(a) We prove $\mathbb{E}(\boldsymbol{b}-\boldsymbol{\beta}|\mathbf{X})=\mathbf{0}$. Note that

$$\boldsymbol{b}-\boldsymbol{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\equiv\mathbf{A}\boldsymbol{\varepsilon}.$$

So

$$\mathbb{E}(\boldsymbol{b}-\boldsymbol{\beta}|\mathbf{X})=\mathbb{E}(\mathbf{A}\boldsymbol{\varepsilon}|\mathbf{X})=\mathbf{A}\mathbb{E}(\boldsymbol{\varepsilon}|\mathbf{X})=\mathbf{0}.$$

(b)

$$\begin{aligned}\mathrm{Var}(\boldsymbol{b}|\mathbf{X}) &= \mathrm{Var}(\boldsymbol{b}-\boldsymbol{\beta}|\mathbf{X})\\ &= \mathrm{Var}(\mathbf{A}\boldsymbol{\varepsilon}|\mathbf{X})\\ &= \mathbf{A}\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X})\mathbf{A}'\\ &= \sigma^2\mathbf{A}\mathbf{A}'\\ &= \sigma^2\cdot(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

□

Proof.

(c) Since $\hat{\boldsymbol{\beta}}$ is linear in $\mathbf{y}$, it can be written as $\hat{\boldsymbol{\beta}} = \mathbf{Cy}$ for some matrix $\mathbf{C}$. Let $\mathbf{D} \equiv \mathbf{C} - \mathbf{A}$ where $\mathbf{A} \equiv (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Then

$$\hat{\boldsymbol{\beta}} = (\mathbf{D} + \mathbf{A})\mathbf{y} = \mathbf{DX}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\varepsilon} + \boldsymbol{b}.$$

Taking the conditional expectation of both sides, we obtain

$$\underbrace{\mathbb{E}(\hat{\boldsymbol{\beta}}|\mathbf{X})}_{\boldsymbol{\beta}} = \mathbf{DX}\boldsymbol{\beta} + \underbrace{\mathbb{E}(\mathbf{D}\boldsymbol{\varepsilon}|\mathbf{X})}_{\mathbf{0}} + \underbrace{\mathbb{E}(\boldsymbol{b}|\mathbf{X})}_{\boldsymbol{\beta}}.$$

It follows that $\mathbf{DX}\boldsymbol{\beta} = \mathbf{0}$. For this to be true for any given $\boldsymbol{\beta}$, it is necessary that $\mathbf{DX} = \mathbf{0}$. So $\hat{\boldsymbol{\beta}} = \mathbf{D}\boldsymbol{\varepsilon} + \boldsymbol{b}$ and

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathbf{D}\boldsymbol{\varepsilon} + (\boldsymbol{b} - \boldsymbol{\beta}) = (\mathbf{D} + \mathbf{A})\boldsymbol{\varepsilon}.$$

Therefore,

$$\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) &= \mathrm{Var}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\mathbf{X}) \\
&= \sigma^2 \cdot (\mathbf{DD}' + \mathbf{AD}' + \mathbf{DA}' + \mathbf{AA}') \\
&\geq \sigma^2 \cdot (\mathbf{X}'\mathbf{X})^{-1} = \mathrm{Var}(\boldsymbol{b}|\mathbf{X}).
\end{aligned}$$

□

Proof.

(d)

$$
\begin{aligned}
\mathrm{Cov}(\boldsymbol{b}, \boldsymbol{e}|\mathbf{X}) &\equiv \mathbb{E}\left\{[\boldsymbol{b} - \mathbb{E}(\boldsymbol{b}|\mathbf{X})][\boldsymbol{e} - \mathbb{E}(\boldsymbol{e}|\mathbf{X})]'|\mathbf{X}\right\} \\
&= \mathbb{E}\left\{\mathbf{A}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{M}|\mathbf{X}\right\} \\
&= \mathbf{A}\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X})\mathbf{M} \\
&= \sigma^2\mathbf{A}\mathbf{M} \\
&= \mathbf{0}.
\end{aligned}
$$

□

σ^2 的估计

The OLS estimate of σ^2, denoted s^2, is the sum of squared residuals divided by $N-K$:

$$s^2 \equiv \frac{\boldsymbol{e}'\boldsymbol{e}}{N-K}.$$

Theorem (Unbiasedness of s^2)

Under Assumptions 1.1-1.4, $\mathbb{E}(s^2|\mathbf{X}) = \sigma^2$, provided $N > K$.

Proof.

Since $s^2 = \boldsymbol{e}'\boldsymbol{e}/(N-K)$, the proof amounts to showing that

$$\mathbb{E}(\boldsymbol{e}'\boldsymbol{e}|\mathbf{X}) = (N-K)\sigma^2$$

Proof.

The proof consists of proving two properties:

(1) $\mathbb{E}(\boldsymbol{e}'\boldsymbol{e}|\mathbf{X}) = \mathbb{E}(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}|\mathbf{X}) = \sigma^2 \cdot \mathrm{trace}(\mathbf{M})$

(2) $\mathrm{trace}(\mathbf{M}) = N - K$.

For (1), since $\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon} = \sum_{i=1}^N \sum_{j=1}^N m_{ij}\varepsilon_i\varepsilon_j$, we have

$$\mathbb{E}(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}|\mathbf{X}) = \sum_{i=1}^N \sum_{j=1}^N m_{ij}\mathbb{E}(\varepsilon_i\varepsilon_j|\mathbf{X}) = \sigma^2 \sum_{i=1}^N m_{ii} = \sigma^2 \cdot \mathrm{trace}(\mathbf{M}).$$

For (2), note that

$$\mathrm{trace}(\mathbf{M}) = \mathrm{trace}(\mathbf{I}_N - \mathbf{P}) = N - \mathrm{trace}(\mathbf{P}),$$

and

$$\begin{aligned} \mathrm{trace}(\mathbf{P}) &= \mathrm{trace}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \\ &= \mathrm{trace}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] \\ &= \mathrm{trace}(\mathbf{I}_K) = K. \end{aligned}$$

So $\mathrm{trace}(\mathbf{M}) = N - K$. □

小样本统计推断

小样本统计推断

5. Assumption 1.5 (normality of the error term): The distribution of ε conditional on $\mathbf{X}$ is jointly normal:

$$\varepsilon \mid \mathbf{X} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N).$$

Remark

- The distribution of ε conditional on $\mathbf{X}$ does not depend on $\mathbf{X}$. It then follows that ε and $\mathbf{X}$ are independent.
- Under Assumptions 1.1-1.5,

$$(\boldsymbol{b} - \boldsymbol{\beta}) \mid \mathbf{X} \sim N(\mathbf{0}, \sigma^2 \cdot (\mathbf{X}'\mathbf{X})^{-1}).$$

单个回归系数的假设检验

Theorem (distribution of the t-ratio)

Suppose Assumptions 1.1–1.5 hold. Under the null hypothesis $\mathbb{H}_0 : \beta_k = \bar{\beta}_k$, the t-ratio defined as

$$t_k \equiv \frac{b_k - \bar{\beta}_k}{SE(b_k)} \equiv \frac{b_k - \bar{\beta}_k}{\sqrt{s^2 \cdot \left((\mathbf{X}'\mathbf{X})^{-1}\right)_{kk}}}$$

is distributed as t_{N-K}.[a]

[a] If $x \sim N(0,1)$, $y \sim \chi^2(m)$ and if x and y are independent, then the ratio $x/\sqrt{y/m}$ has the t distribution with m degrees of freedom.

Proof.

We can write

$$t_k = \frac{b_k - \bar{\beta}_k}{\sqrt{\sigma^2 \cdot ((\mathbf{X}'\mathbf{X})^{-1})_{kk}}} \cdot \sqrt{\frac{\sigma^2}{s^2}} \equiv \frac{z_k}{\sqrt{s^2/\sigma^2}} = \frac{z_k}{\sqrt{\frac{\boldsymbol{e}'\boldsymbol{e}/(N-K)}{\sigma^2}}} \equiv \frac{z_k}{\sqrt{\frac{q}{N-K}}},$$

where $q = \boldsymbol{e}'\boldsymbol{e}/\sigma^2$. Note that z_k is $N(0,1)$. We will show:

(1) $q \mid \mathbf{X} \sim \chi^2(N-K)$,

(2) z_k and q are independent conditional on $\mathbf{X}$.

First, note that

$$q = \frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2} = \frac{\boldsymbol{\varepsilon}'}{\sigma}\mathbf{M}\frac{\boldsymbol{\varepsilon}}{\sigma}.$$

Since $\mathbf{M}$ is idempotent and $\boldsymbol{\varepsilon}/\sigma \mid \mathbf{X} \sim N(\mathbf{0}, \mathbf{I}_N)$, we have $q \mid \mathbf{X} \sim \chi^2(\mathrm{rank}(\mathbf{M}))$ with $\mathrm{rank}(\mathbf{M}) = N - K$. Lemma

For (2), both $\boldsymbol{b}$ and $\boldsymbol{e}$ are linear functions of $\boldsymbol{\varepsilon}$, so they are jointly normal conditional on $\mathbf{X}$. Also, they are uncorrelated conditional on $\mathbf{X}$. So $\boldsymbol{b}$ and $\boldsymbol{e}$ are independently distributed conditional on $\mathbf{X}$. But z_k is a function of $\boldsymbol{b}$ and q is a function of $\boldsymbol{e}$. So z_k and q are independently distributed conditional on $\mathbf{X}$. $\square$

Back

Lemma

If $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_N)$ *and* $\mathbf{A}$ *is symmetric and idempotent, then* $\mathbf{x}'\mathbf{A}\mathbf{x}$ *has a chi-squared distribution with degrees of freedom equal to the rank of* $\mathbf{A}$*.*

Proof.

Let the characteristic roots and vectors of $\mathbf{A}$ be arranged in a diagonal matrix $\mathbf{\Lambda}$ and an orthogonal matrix $\mathbf{C}$. Then

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{C}\mathbf{\Lambda}\mathbf{C}'\mathbf{x}$$

where $\mathbf{C}$ satisfies the requirement that $\mathbf{C}'\mathbf{C} = \mathbf{I}_N$. Note that $\mathbf{y} \equiv \mathbf{C}'\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_N)$. So

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{N} \lambda_i y_i^2 = \sum_{i=1}^{J} y_i^2 \sim \chi^2(J),$$

where $J = \operatorname{rank}(\mathbf{A})$. □

Back

Lemma

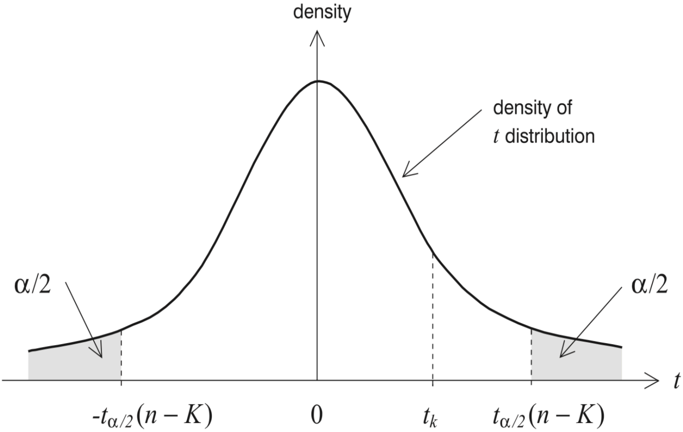
If $\mathbf{A}$ *is idempotent, then* $\operatorname{rank}(\mathbf{A}) = \operatorname{trace}(\mathbf{A})$.

Proof.

See [Abadir & Magnus, 2005], Exercise 8.61. □

Note that

- The trace of a matrix equals the sum of its characteristic roots.
- The rank of a symmetric matrix is the number of nonzero characteristic roots it contains.
- The characteristic roots of an idempotent matrix are all zero or one.


density
density of
t distribution
α/2
α/2
t
-tα/2(n − K)
0
tk
tα/2(n − K)

一般线性假设检验

Theorem (distribution of the F-ratio)

Suppose Assumptions 1.1–1.5 hold. Under the null hypothesis $\mathbb{H}_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, *where* J *is the dimension of* $\mathbf{r}$ *and* $\mathbf{R}$ *is* $J \times K$ *with* $\operatorname{rank}(\mathbf{R}) = J$, *the* F*-ratio defined as*

$$\begin{aligned} F &\equiv \frac{(\mathbf{R}\boldsymbol{b} - \mathbf{r})' \left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1} (\mathbf{R}\boldsymbol{b} - \mathbf{r})/J}{s^2} \\ &= (\mathbf{R}\boldsymbol{b} - \mathbf{r})' \left[\mathbf{R}\widehat{\operatorname{Var}(\boldsymbol{b}|\mathbf{X})}\mathbf{R}'\right]^{-1} (\mathbf{R}\boldsymbol{b} - \mathbf{r})/J \end{aligned}$$

is distributed as $F(J, N - K)$.[a]

[a] If x_1 and x_2 are two independent chi-squared variables with degrees of freedom n_1 and n_2 respectively, then the ratio $(x_1/n_1)/(x_2/n_2)$ has the F distribution with n_1 and n_2 degrees of freedom.

Proof.

Since $s^2 = \boldsymbol{e}'\boldsymbol{e}/(N-K)$, we can write

$$F = \frac{w/J}{q/(N-K)}$$

where

$$w \equiv (\mathbf{R}\boldsymbol{b} - \mathbf{r})' \left[\sigma^2 \cdot \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1} (\mathbf{R}\boldsymbol{b} - \mathbf{r}) \quad \text{and} \quad q \equiv \frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}.$$

We need to show

(1) $w \mid \mathbf{X} \sim \chi^2(J)$,

(2) $q \mid \mathbf{X} \sim \chi^2(N-K)$,

(3) w and q are independently distributed conditional on $\mathbf{X}$.

For (1), let $\mathbf{v} \equiv \mathbf{R}\boldsymbol{b} - \mathbf{r}$. Under $\mathbb{H}_0$, $\mathbf{v} = \mathbf{R}(\boldsymbol{b} - \boldsymbol{\beta})$. So, $\mathbf{v} \mid \mathbf{X} \sim N(\mathbf{0}, \sigma^2 \cdot \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')$. Hence, $w \mid \mathbf{X} \sim \chi^2(J)$. ▸ Lemma

To establish (3), note that w is a function of $\boldsymbol{b}$ and q is a function of $\boldsymbol{e}$. □

Back

Lemma

Let $\mathbf{x}$ *be an* m *dimensional random vector. If* $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *with* $\boldsymbol{\Sigma}$ *nonsingular, then* $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \sim \chi^2(m)$.

Proof.

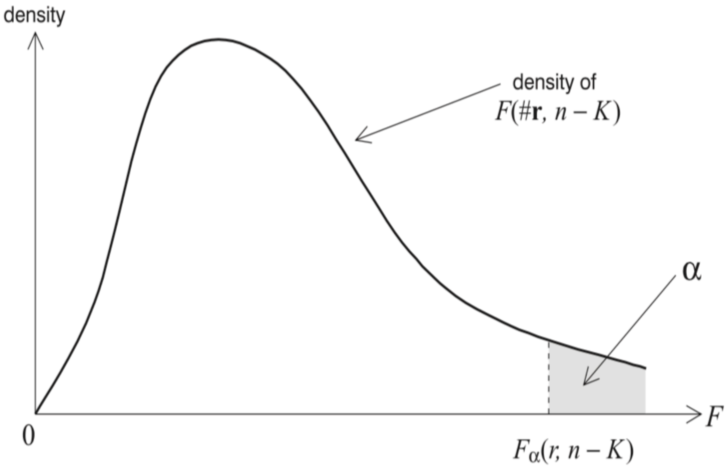
Because $\boldsymbol{\Sigma}$ is symmetric and positive definite, we can find a symmetric and invertible matrix $\boldsymbol{\Sigma}^{1/2}$ such that

$$
\begin{aligned}
\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} &= \boldsymbol{\Sigma}, \\
\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} &= \boldsymbol{\Sigma}^{-1}.
\end{aligned}
$$

Now, define $\mathbf{y} \equiv \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \mu)$. It follows that $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I}_m)$. Therefore, we have

$$\mathbf{y}'\mathbf{y} \sim \chi^2(m).$$

□


density
density of
$F(\#\mathbf{r}, n-K)$
α
F
0
$F_\alpha(r, n-K)$

F 统计量的另一种表达式

F 统计量还可以表示为：

$$F = \frac{(SSR_R - SSR_U)/J}{SSR_U/(N-K)},$$

其中，SSR_U 是无约束的残差平方和，SSR_R 是有约束的残差平方和，由求解如下最优化问题得到：

$$\min_{\tilde{\boldsymbol{\beta}}} SSR(\tilde{\boldsymbol{\beta}}) \quad \text{s.t.} \quad \mathbf{R}\tilde{\boldsymbol{\beta}} = \mathbf{r}.$$

Proof.

Let $\tilde{\boldsymbol{b}}$ be the OLS under $\mathbb{H}_0$, that is,

$$\tilde{\boldsymbol{b}} = \arg\min_{\tilde{\boldsymbol{\beta}}} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \quad \text{s.t.} \quad \mathbf{R}\tilde{\boldsymbol{\beta}} = \mathbf{r}.$$

We first form the Lagrangian function

$$L(\tilde{\boldsymbol{\beta}}, \boldsymbol{\lambda}) = (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) + 2\boldsymbol{\lambda}'(\mathbf{r} - \mathbf{R}\tilde{\boldsymbol{\beta}}),$$

where $\boldsymbol{\lambda}$ is a $J \times 1$ vector called the Lagrange multiplier vector.

Proof.

We have the following FOC:

$$\begin{aligned}\frac{\partial L(\tilde{\boldsymbol{b}}, \tilde{\boldsymbol{\lambda}})}{\partial \tilde{\boldsymbol{\beta}}} &= -2\mathbf{X}'(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{b}}) - 2\mathbf{R}'\tilde{\boldsymbol{\lambda}} = \mathbf{0}, \\ \frac{\partial L(\tilde{\boldsymbol{b}}, \tilde{\boldsymbol{\lambda}})}{\partial \boldsymbol{\lambda}} &= 2(\mathbf{r}-\mathbf{R}\tilde{\boldsymbol{b}}) = \mathbf{0}.\end{aligned}$$

From the first equation of FOC, we can obtain

$$-\mathbf{R}(\boldsymbol{b}-\tilde{\boldsymbol{b}}) = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\tilde{\boldsymbol{\lambda}}.$$

Hence,

$$\tilde{\boldsymbol{\lambda}} = -[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\boldsymbol{b}-\mathbf{r}).$$

It follows that

$$\boldsymbol{b}-\tilde{\boldsymbol{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\boldsymbol{b}-\mathbf{r}).$$

Now,

$$\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{b}} = \mathbf{y}-\mathbf{X}\boldsymbol{b}+\mathbf{X}(\boldsymbol{b}-\tilde{\boldsymbol{b}}) = \boldsymbol{e}+\mathbf{X}(\boldsymbol{b}-\tilde{\boldsymbol{b}}).$$

It follows that

Proof.

$$
\begin{aligned}
SSR_R &= (\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{b}})'(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{b}}) \\
&= \boldsymbol{e}'\boldsymbol{e} + (\boldsymbol{b}-\tilde{\boldsymbol{b}})'\mathbf{X}'\mathbf{X}(\boldsymbol{b}-\tilde{\boldsymbol{b}}) \\
&= \boldsymbol{e}'\boldsymbol{e} + (\mathbf{R}\boldsymbol{b}-\mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\boldsymbol{b}-\mathbf{r}).
\end{aligned}
$$

We have

$$
(\mathbf{R}\boldsymbol{b}-\mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\boldsymbol{b}-\mathbf{r}) = SSR_R - SSR_U
$$

and

$$
\begin{aligned}
F &= \frac{(\mathbf{R}\boldsymbol{b}-\mathbf{r})'\left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}(\mathbf{R}\boldsymbol{b}-\mathbf{r})/J}{s^2} \\
&= \frac{(SSR_R - SSR_U)/J}{SSR_U/(N-K)}.
\end{aligned}
$$

□

特例：解释变量联合显著性检验

考虑如下线性回归模型：

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i,$$

和原假设 $\mathbb{H}_0 : \beta_j = 0$ for $2 \le j \le K$. 容易证明，有约束的残差平方和为

$$SSR_R = \sum_{i=1}^{N} (y_i - \bar{y})^2.$$

由 R^2 的定义，我们有：

$$R^2 = 1 - \frac{\boldsymbol{e}'\boldsymbol{e}}{SSR_R}.$$

于是，

$$\begin{aligned} F &= \frac{(SSR_R - \boldsymbol{e}'\boldsymbol{e})/(K-1)}{\boldsymbol{e}'\boldsymbol{e}/(N-K)} = \frac{\left(1 - \dfrac{\boldsymbol{e}'\boldsymbol{e}}{SSR_R}\right)/(K-1)}{\dfrac{\boldsymbol{e}'\boldsymbol{e}}{SSR_R}/(N-K)} \\ &= \frac{R^2/(K-1)}{(1-R^2)/(N-K)}. \end{aligned}$$

References

Fumio Hayashi. (2000)

Econometrics, Chapter 1.

Princeton University Press, 2000.

Karim M. Abadir & Jan R. Magnus. (2005)

Matrix Algebra.

Cambridge University Press, 2005.